applied sciences

MDPI

Article

WDTISeg: One-Stage Interactive Segmentation for Breast Ultrasound Image Using Weighted Distance Transform and Shape-Aware Compound Loss

Xiaokang Li [1], Mengyun Qiao [1], Yi Guo [1,*], Jin Zhou [2], Shichong Zhou [2], Cai Chang [2] and Yuanyuan Wang [1,*]

[1] Department of Electronic Engineering, Fudan University, Shanghai 200433, China; lixiaokang@fudan.edu.cn (X.L.); myqiao7@fudan.edu.cn (M.Q.)
[2] Fudan University Shanghai Cancer Center, Shanghai 200032, China; 19111230050@fudan.edu.cn (J.Z.); sczhou@fudan.edu.cn (S.Z.); changcai@fudan.edu.cn (C.C.)
* Correspondence: guoyi@fudan.edu.cn (Y.G.); yywang@fudan.edu.cn (Y.W.)

Abstract: Accurate tumor segmentation is important for aided diagnosis using breast ultrasound. Interactive segmentation methods can obtain highly accurate results by continuously optimizing the segmentation result via user interactions. However, traditional interactive segmentation methods usually require a large number of interactions to make the result meet the requirements due to the performance limitations of the underlying model. With greater ability in extracting image information, convolutional neural network (CNN)-based interactive segmentation methods have been shown to effectively reduce the number of user interactions. In this paper, we proposed a one-stage interactive segmentation framework (interactive segmentation using weighted distance transform, WDTISeg) for breast ultrasound image using weighted distance transform and shape-aware compound loss. First, we used a pre-trained CNN to attain an initial automatic segmentation, based on which the user provided interaction points of mis-segmented areas. Then, we combined Euclidean distance transform and geodesic distance transform to convert interaction points into weighted distance maps to transfer segmentation guidance information to the model. The same CNN accepted the input image, the initial segmentation, and weighted distance maps as a concatenation input and provided a refined result, without another additional segmentation network. In addition, a shape-aware compound loss function using prior knowledge was designed to reduce the number of user interactions. In the testing phase on 200 cases, our method achieved a dice of 82.86 ± 16.22 (%) for automatic segmentation task and a dice of 94.45 ± 3.26 (%) for interactive segmentation task after 8 interactions. The results of comparative experiments proved that our method could obtain higher accuracy with fewer simple interactions than other interactive segmentation methods.

Keywords: interactive image segmentation; breast ultrasound; weighted distance transform; prior knowledge

check for updates

Citation: Li, X.; Qiao, M.; Guo, Y.; Zhou, J.; Zhou, S.; Chang, C.; Wang, Y. WDTISeg: One-Stage Interactive Segmentation for Breast Ultrasound Image Using Weighted Distance Transform and Shape-Aware Compound Loss. *Appl. Sci.* **2021**, *11*, 6279. https://doi.org/10.3390/app11146279

Academic Editor: Alessandro Ramalli

Received: 7 June 2021
Accepted: 4 July 2021
Published: 7 July 2021

1. Introduction

Breast cancer is one of the leading causes of death in women around the world and diagnosing breast cancer in its early stages will always remain crucial [1,2]. Breast ultrasound is widely used in clinical diagnosis for its advantages of safety and low cost. Generally, accurate tumor segmentation is necessary and significant for precise diagnosis using breast ultrasound. However, fully automatic segmentation methods are difficult to obtain accurate results that can meet clinical analysis standards [3]. This is mainly related to the poor quality of the ultrasound images, but also to the limitations of the segmentation model. Compared to automatic segmentation methods that gives results at once, the advantage of interactive segmentation is that the user provides prior knowledge about the object through interactions to guide the refinement of the segmentation result [4]. In a real clinical situation, each patient may have multiple ultrasound images, and it is unrealistic to use manual annotation of tumor boundaries for all of them. Therefore, interactive segmentation tools with fast implementation of high accuracy segmentation have a significant meaning for clinical use.

There are three key points of excellent interaction segmentation of medical images: simple type of interactions, efficient interaction information transfer, and the use of prior knowledge. Existing interactive segmentation methods have different types of interactions, which can be divided into providing points [5–9], scribbles [10–16], a bounding box (BB), or a polygon box (PB) [11,17]. Among these ways, providing scribbles, BB and PB require the user to swipe the mouse pointer over the image for a long time, while clicking on points is intuitively the easiest interaction type. Interaction information transfer refers to the way of using user interactions to guide the segmentation. Most of the existing interaction segmentation methods use Gaussian probability maps, distance transforms, etc., to transfer user interaction information to the segmentation mode. However, they cannot utilize both the location information of interaction points and the contextual information of the image. Since human interaction is actually providing prior information to the network, using prior information in the model can reduce the number of user interactions but few ways take advantage of this.

Some conventional interactive segmentation methods are based on graph theory. The graph cuts [10] method uses the Gaussian mixture model (GMM) as the underpinning model and needs the user’s scribbles for refinement. In this method, a large number of scribbles are needed before getting a satisfactory accuracy. GrabCut [11] requires the user to provide a bounding box to limit the region of interest (ROI) to take less scribbles, but its performance is poor on medical images as graph cuts [10] on account of the same GMM model. The Random walker segmentation method [12] uses random walker as the basic model to attain a refined result. These three methods all require a lot of user interactions due to the poor performance of underpinning model. In 2007, Bai et al. proposed an interactive framework [13] using geodesic distances to convert user-provided scribbles, so that the target could be automatically segmented. This was the first method to use geodesic distance transform for interactive segmentation, while some subsequent methods [14–16] have improved on this. However, all of them only perform well on images with large differences between foreground and background, because the geodesic distance focuses on the gradient information of the original image.

In order to break through the limitations of traditional method, interactive segmentation methods based on CNNs have been proposed. Xu et al. [5] converted user’s interaction points into Euclidean distance maps based on foreground and background points. The five-channel image (original RGB channels and two distance transform map) was used as the input of a full convolutional network (FCN) to obtain the segmentation result. Euclidean distance transform is concerned with the location information of interaction points and cannot utilize image contexts information. BIFSeg [6] uses an image segmentation method similar to GrabCut [11]. The user first draws a boundary box as the input for CNN to obtain an initial result. Then, image-specific fine-tuning conducts CNN to improve segmentation results. DeepIGeoS [7] firstly proposes using geodesic distance maps as part of the input for CNNs. Geodesic distance maps can reflect the grayscale texture information of the original image by calculating the shortest distance from the full image to a specific point, so that CNNs can identify mis-segmentations of foreground and background from the input data to refine the segmentation result. However, it is sensitive to the contrast and spatial information of the image, and lacks the importance of clearly indicating the location information of the interaction point. For example, in the case of blurred tumor boundary, the geodesic distance near the boundary does not change significantly due to the small image gray gradient change, while the Euclidean distance is only related to the locations of interactions and not influenced by the quality of the original image. This means that the Euclidean distance is more effective than the geodesic distance in pointing out the misalignment area. Therefore, there is an urgent need for a method that combines the advantages of both distance transforms.

In this paper, we proposed a one-stage interactive segmentation framework for breast ultrasound image based on the above three key points. Compared with existing two-stage interactive segmentation networks [6,7] to refine the result of automatic segmentation network, our method has several advantages. First, our method can use the same CNN network (I-net) to obtain the automatic segmentation and refined results in turn. We trained I-net on automatic segmentation task to ensure that it could provide an initial segmentation when inputting the only original image. Second, our method has more effective interaction information transfer. We proposed a weighted distance transform combined geodesic distance and Euclidean distance transforms, which means the distance map could reflect both the texture information near the object area and the location information of the interaction points in the whole image. Third, our method can reduce the number of interactions for the use of prior information in the training phase. We referred to the proposed framework as the interactive segmentation using weighted distance transform (WDTISeg).

The main contributions of the proposed method are as follows:

(1) We proposed a one-stage interactive segmentation framework for breast ultrasound image segmentation, which is the first method to use a network to get both automatic and interactive segmentation. The training process was greatly simplified because no additional automatic segmentation network was required to provide the initial results;

(2) We proposed to convert user interactions into maps with weighted distance transform which combines geodesic distance and Euclidean distance transforms. This combination can effectively convey both location information of interaction points and exploit image contexts knowledge;

(3) We proposed a shape-aware compound loss function using the prior knowledge of breast tumors in the training phase to reduce the number of interactions. The compound loss function improved the accuracy of model segmentation while avoiding oscillation and overfitting in the training process.

2. Methods

2.1. Proposed Framework

Figure 1 shows the pipeline of the proposed framework WDTISeg for interactive segmentation. I-net is the backbone segmentation network of the framework with input data of four channels. It was pre-trained for automatic segmentation task.

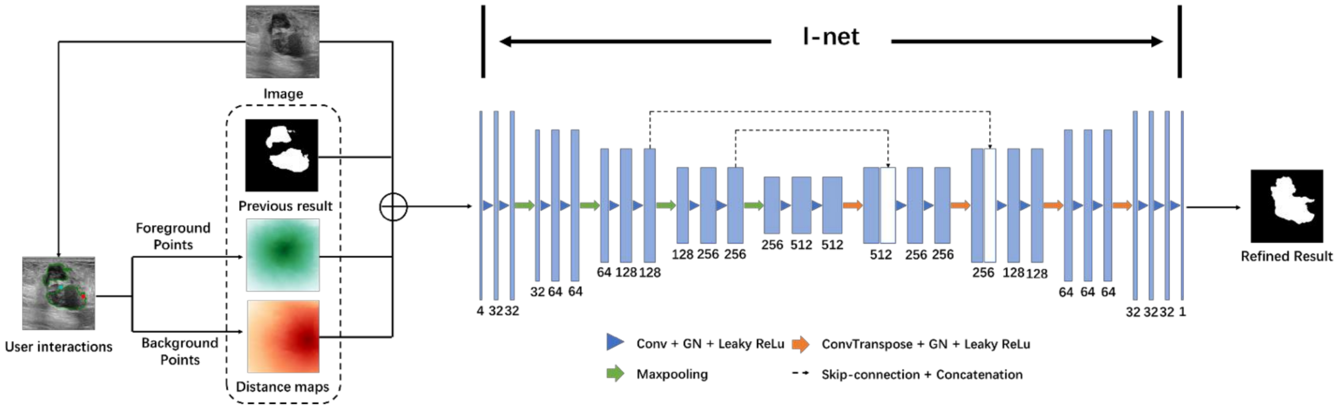

Figure 1. Framework of the proposed method WDTISeg. The cyan dots and red dots in the segmentation represent foreground points and background points, respectively. The number below each block is the number of feature maps.

Firstly, the input image was put into I-net to obtain an initial automatic segmentation, while the other three channels were all zero. Secondly, the user provided interaction points of foreground and background on mis-segmented regions according to the initial segmentation. Then these interaction points were converted into the foreground distance map (cyan) and the background distance map (red) by weighted distance transform, as shown in Figure 1. Finally, the input image, the initial segmentation, and two distance maps were concatenated before being put into I-net to attain a refined segmentation. This interactive process was repeated until we attained a segmentation with satisfactory accuracy.

2.2. The Structure of I-Net

It should be noted that I-net attains automatic segmentation and interactive segmentation depending on the form of the input data. When the image is first fed into the network, neither the initial segmentation nor distance maps exist, so I-net provides an output of automatic segmentation. Compared with two-stage interactive segmentation methods or refined segmentation methods, such as DeepIGeoS [7], our method does not need an additional CNN to obtain an initial segmentation, making our model lighter and easy to train.

Figure 1 also shows details of I-net in our method. The network received a four-channel data as input to predict the segmentation result. As shown in Figure 1, I-net is designed based on U-net [18] with an encoder–decoder architecture.

I-net improved U-net [18] in several parts to fit our segmentation tasks. We used group normalization [19] to replace batch normalization [20] to make normalization free from the dependence on batch size. In addition, we used the leaky ReLu [21] layer instead of the ReLu [22] layer to solve the problem of dead neuron while retaining the advantages of ReLu function. Leaky ReLu is defined as follows:

$$\text{Leaky ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0 \end{cases}, \tag{1}$$

where the α was set to 0.2 in this work.

To avoid the noise in the input image introduced by the shallow layers' skip connection to affect the segmentation results, I-net retained only two middle skip connections compared to U-net. These two skip connections aimed to utilize the combination of low-level features and high-level information to achieve better segmentation of tumor margins. At the last stage of the decoder, a convolution layer with one filer was used to attain the final segmentation.

2.3. Weighted Distance Transform

We proposed a weighted distance transform to convert user interactions into distance maps. An image can be regarded as an undirected graph with weight. Each pixel is a node and the grayscale difference between neighboring pixels is the weights of the edge. In graph theory, the geodesic distance is the distance of the shortest path between two nodes in a graph, so the geodesic distance map can reflect the grayscale texture information of the original image. The Euclidean distance is the shortest distance between two points in geometric space.

Let i and j be two different pixels in an image I, then the unsigned geodesic distance between i and j is:

$$D_{Geo}(i,\ j,\ \boldsymbol{I}) = \min_{p \in \mathcal{P}_{i,\ j}} \int_0^1 \parallel \nabla \boldsymbol{I}(p(s) \cdot \boldsymbol{u}(s)) \parallel ds, \tag{2}$$

where $\mathcal{P}_{i,\ j}$ is the set of all paths between pixel i and j. $p(s)$ is one feasible path and it is parameterized by s $\in$ [0, 1]. $\boldsymbol{u}(s)$ is a unit vector that is tangent to the direction of $p(s)$ [7,13,14].

The unsigned Euclidean distance between i and j is:

$$D_{Euc}(i,\ j,\ \boldsymbol{I}) = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}, \tag{3}$$

When we combine the two distances, let the weighted distance between pixel *i* and *j* be:

$$D_w(i,\ j,\ \boldsymbol{I}) = \lambda \cdot D_{Geo}(i,\ j,\ \boldsymbol{I}) + (1-\lambda) \cdot D_{Euc}(i,\ j,\ \boldsymbol{I}), \tag{4}$$

where λ is a hyperparameter that requires experimental verification. Weighted distance turns into Euclidean distance when $\lambda = 0$, and geodesic distance when $\lambda = 1$, respectively.

Suppose S_f and S_b represent the foreground points set and the background points set, respectively. Then the unsigned weighted distance from *i* to the points set *S* ($S \in \left\{S_f,\ S_b\right\}$ is:

$$G(i,\ S,\ \boldsymbol{I}) = \min_{j \in S} D_w(i,\ j,\ \boldsymbol{I}), \tag{5}$$

There are already many algorithms in computer science for solving optimization Equation (3), such as the Floyd's algorithm, the Dijkstra's algorithm [23], and the fast marching algorithm [24]. Here we use fast marching for its speediness. The geodesic distance map was set to all zeros if no points in the foreground region and background region were clicked. Figure 2 shows an example of weighted distance maps when λ in (4) takes different values.

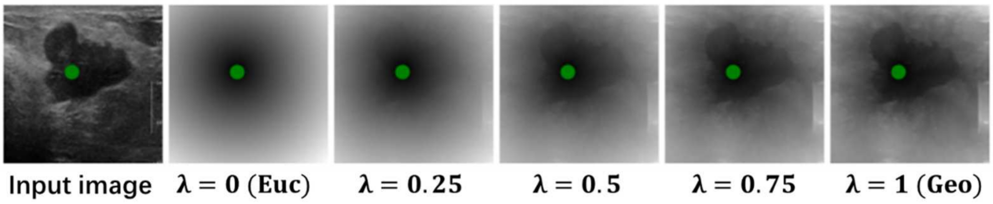

Figure 2. An example of weighted distance maps when λ in (3) takes different values. The green point is the foreground interaction point.

2.4. Training and Testing of I-Net

For fast and efficient model training, we used automatically generated simulated interaction points in the model training phase, while interaction points were obtained by user clicks on images in the testing phase.

2.4.1. Training

In the training phase, in order to quickly and automatically build the model for interaction segmentation, we generated interaction points that simulated the user's clicks by comparing the ground truth (f_y) with the initial segmentation (f_x). The subtraction of the two images could provide a mis-segmented foreground region and background region, as shown in (6).

$$f_z = f_x - f_y \begin{cases} \text{background region}, & f_z < 0 \\ \text{foreground region}, & f_z > 0 \end{cases}, \tag{6}$$

Then user interaction points can be automatically generated from each mis-segmentation region by randomly sampling n pixels in that region. The number of pixels of the region is *N*. In this work, n was determined as the follow function (7) by experience.

$$\text{n} = \begin{cases} 0, & N < 100 \\ ceil\left(\frac{N}{500}\right), & \text{other} \end{cases}, \tag{7}$$

where $ceil(x)$ returns the smallest integer value greater than or equal to x.

Figure 3 shows some examples of simulated interaction points in the training phase.

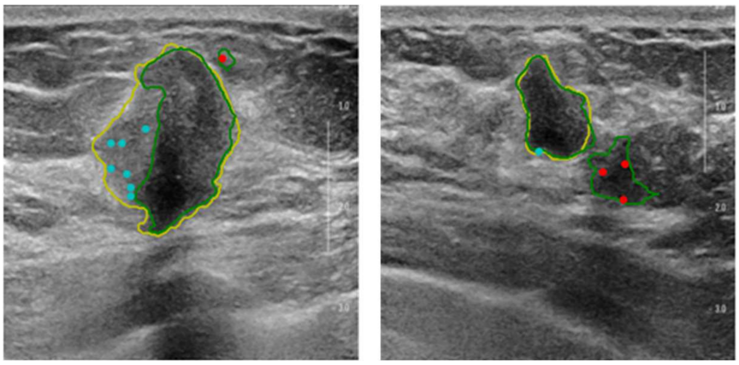

Figure 3. Examples of simulated interaction points in the training phase. The yellow and green lines are the contours of the ground truth and the previous segmentation, respectively. The cyan and red points are foreground and background interaction points, respectively.

2.4.2. Testing

Interaction points in the testing phase are obtained by the operator by clicking on the mis-segmented region as shown in Figure 4. Instead of having a ground truth in the training phase, the user clicks with points on mis-segmented areas with prior knowledge. In each interaction phase, the user should give one foreground point and one background point with reference to the initial segmentation from P-net or the segmentation result of previous interaction.

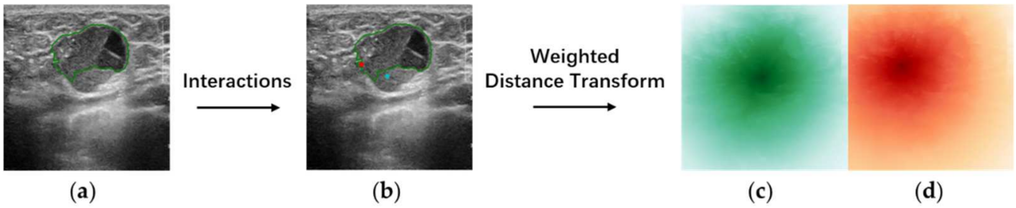

Figure 4. The process of getting interaction points in the testing phase. (**a**) The input image with the initial segmentation (green line); (**b**) the foreground point (cyan) and background point (red) given by the user interaction; (**c**) the foreground point distance map; (**d**) the background point distance map.

As shown in the framework shown in Figure 1, the interaction process continues until the user is satisfied with the result of the segmentation or the maximum threshold of interactions is reached, which was set to 8 in our study.

2.5. Loss Function

Incorporating prior knowledge into the loss function is important to improve segmentation accuracy and reduce the number of user interactions [25]. By observing breast tumors in ultrasound images, we found some prior knowledge that is useful for tumor segmentation. First, most tumors are actually compact contiguous domains. Some benign tumors are even closer to round or ellipse shapes. Second, the physician usually ensures that only one tumor remains on the image when saving the breast ultrasound. Even when there are two or more tumors on the image, they can be separated by cropping the image.

Based on the above findings, we proposed a shape-aware compound loss function $\mathcal{L}_{\text{total}}$ to incorporate prior knowledge with CNN. As defined in (8), $\mathcal{L}_{\text{total}}$ is composed by binary cross entropy loss ($\mathcal{L}_{\text{BCE}}$), dice loss ($\mathcal{L}_{\text{Dice}}$), and shape constraint loss ($\mathcal{L}_{\text{SC}}$).

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} - \log(1 - \mathcal{L}_{Dice}) + \omega \cdot \mathcal{L}_{\text{SC}}, \tag{8}$$

where ω is the weight of $\mathcal{L}_{\text{SC}}$.

Here $\mathcal{L}_{\text{SC}}$ is the loss function we used for the shape constraint:

$$\mathcal{L}_{\text{SC}} = \frac{P^2}{4\pi S}, \tag{9}$$

where S and P are the area and the perimeter of the prediction segmentation.

When the predicted tumor shape is circular, the loss $\mathcal{L}_{\text{SC}}$ is the minimum value of 1. Since the tumor area is compact and connected, the loss $\mathcal{L}_{\text{SC}}$ is be greater than 1 when multiple areas are segmented or the tumor shape is dispersed. Since the shape constraint itself converges to 1 when the shape is a circle, in order to use only its compact shape constraint function without affecting the overall segmentation effect, ω takes 0.05 here.

Cross entropy (CE) is commonly used as a loss function in deep learning and binary cross entropy (BCE) can be used as a loss function in binary classification tasks. The formula for BCE is as follows:

$$\mathcal{L}_{\text{BCE}} = -Y\log(X) - (1-Y)\log(1-X), \tag{10}$$

where X and Y represent the segmentation of the method and the ground truth, respectively.

Dice loss is a dice-based loss function. The reason why dice loss is sometimes directly used as the loss function is that the real goal of segmentation is to maximize the dice coefficient. In general, the use of dice loss has a negative impact on back propagation and tends to make the training unstable.

$$\mathcal{L}_{\text{Dice}} = 1 - \text{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \tag{11}$$

where X and Y represent the same things with (10).

3. Experiments Results and Discussion

3.1. Setting

A dataset of 2200 breast ultrasound images was acquired in Fudan University Shanghai Cancer Center, Shanghai, China from January 2019 to December 2019. The equipment used to obtain ultrasound images included the Aixplorer ultrasound system (SuperSonic Imagine S.A., Aix-en-Provence, France) at 7–15 MHz and the Resona 5S ultrasound system (Shenzhen Mindray Bio-Medical Electronics Co. Ltd., Shenzhen, China) at 5–14 MHz. All images were stored in DICOM format. Each ultrasound image has a tumor segmentation that has been precisely outlined by an experienced radiologist as the ground truth. The image size range is from 721 × 496 to 931 × 606. All images are resized to 256 × 256 before being fed into the network.

All images are arranged in chronological order of the patients' diagnosis. We used the first 2000 cases for training and the remaining 200 cases for testing, which ensured the independence of the patients in our training dataset and testing dataset.

For the quantitative evaluation, our work employed the dice value (dice) (%).

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{12}$$

where X and Y represent the same qualities as in (10).

3.2. Implementation Details This Is Example 1 of an Equation

Adam [26] with a learning rate at 3×10^{-4} was used to be the optimizer in the training stage. The batch size was 32 and the ratio of validation was set to 20% (200 cases). The model was trained for 50 epochs and only saved at the best validation loss. We trained and tested our interactive network using an Intel(R) Xeon(R) Gold 6130 CPU at 2.10 GHz and an NVIDIA TESLA V100 (32G).

Our WDTIseg was at low cost during the training and testing phases. In the training phase, WDTIseg was trained with different λ and loss functions, while the average training time was 624.6 s. The model size was 385 Mb. In the testing phase, the time from the input image put into the network to attain the final segmentation after 8 interactions was recorded, while the average cost was 17.6 s.

3.3. Performance on Automatic Segmentation Task

Our proposed framework WDTISeg could both obtain automatic segmentation and refine results based on interactions. To demonstrate that our method did not require an additional training of an automatic segmentation network to obtain the initial segmentation, we compared the automatic segmentation results of U-net and WDTISeg.

Table 1 shows automatic segmentation results of U-net and WDTISeg. The dice of automatic segmentation results of WDTISeg was 82.86 ± 16.22 (%), better than that of U-net. In the automatic segmentation examples in Figure 5, it is clear that the results of WDTISeg were similar to U-net, and the segmentation results were even slightly more compact.

Table 1. A comparison of dice values of the final segmentation results after 8 interactions. Automatic segmentation results of U-net and WDTISeg are presented to reflect the gain of the interactive segmentation methods.

Methods	Dice (%)
U-net (automatic)	80.50 ± 18.78
WDTISeg (automatic)	82.86 ± 16.22
Graph cuts	82.46 ± 15.53
Random walker	85.16 ± 13.85
DeepIGeoS (R-net)	92.49 ± 7.05
WDTISeg ($\lambda = 0$, $\mathcal{L}_{\text{total}}$)	92.28 ± 7.13
WDTISeg ($\lambda = 0.25$, $\mathcal{L}_{\text{total}}$)	93.89 ± 4.70
WDTISeg ($\lambda = 0.5$, $\mathcal{L}_{\text{total}}$)	94.45 ± 3.26
WDTISeg ($\lambda = 0.75$, $\mathcal{L}_{\text{total}}$)	93.54 ± 3.63
WDTISeg ($\lambda = 1$, $\mathcal{L}_{\text{total}}$)	92.87 ± 6.09
WDTISeg ($\lambda = 0.5$, $\mathcal{L}_{\text{Dice}}$)	92.17 ± 7.29
WDTISeg ($\lambda = 0.5$, $\mathcal{L}_{\text{BCE}}$)	93.01 ± 6.46
WDTISeg ($\lambda = 0.5$, $\mathcal{L}_{\text{Dice+BCE}}$)	93.54 ± 3.63

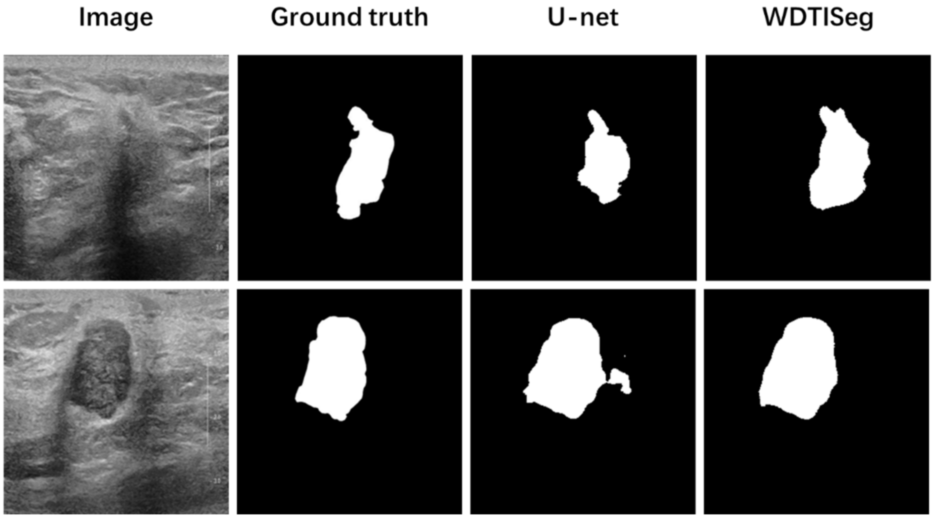

Figure 5. The automatic segmentation results of U-net and WDTISeg.

These prove that WDTISeg can still have comparable automatic segmentation performance to U-net after interactive segmentation training.

Our study focused on improving automatic segmentation-based refinement, and for the first time, we proposed that the interactive segmentation network can generate the initial segmentation results by itself without the need to train additional automatic segmentation networks.

3.4. Impact of the Factor λ in Weighted Distance Transform

To verify the effectiveness of combining the two distance transforms, we compared the single interaction results when λ took different values. Different values represent the different weights of the two distance transforms. The weighted distance became Euclidean distance completely when λ took 0, and Geodesic distance completely when λ took 1. However, we used the same user interactions during the experiment.

As can be seen from Table 1, the interactive segmentation method performed much better than the conventional automatic segmentation method U-net, by as much as 10%. Our method with the parameter $\lambda = 0.5$, $\mathcal{L}_{\text{total}}$ achieved a dice score of $94.45 \pm 3.26\%$ and it performed better than the other four values of λ. By fixing the loss function to be $\mathcal{L}_{\text{total}}$, we can see that the results when λ was t between 0 and 1 were better than both 0 and 1. This proved that combining the two distance conversions can perform better than using either method alone on an interactive segmentation task.

Figure 6 shows a comparison of the segmentation results of our method with different values of λ by given the same user interaction points. The upper case 1 is a tumor with an obscure border, where the interaction point location information is more important than the texture information. In this case, the performance of Euclidean distance transform should be better than Geodesic transform, which is as the same in Figure 6. In the lower case 2, the tumor boundary is obvious, but it has a mis-segmentation outside the tumor. This requires both texture information to ensure correct segmentation of the tumor region and interaction point location information to instruct the network to remove mis-segmented regions outside the tumor. Therefore, λ of 0.5 is better than any other value in case 2. This proves that the combination of our two distance transforms is beneficial in dealing with tumors in different cases.

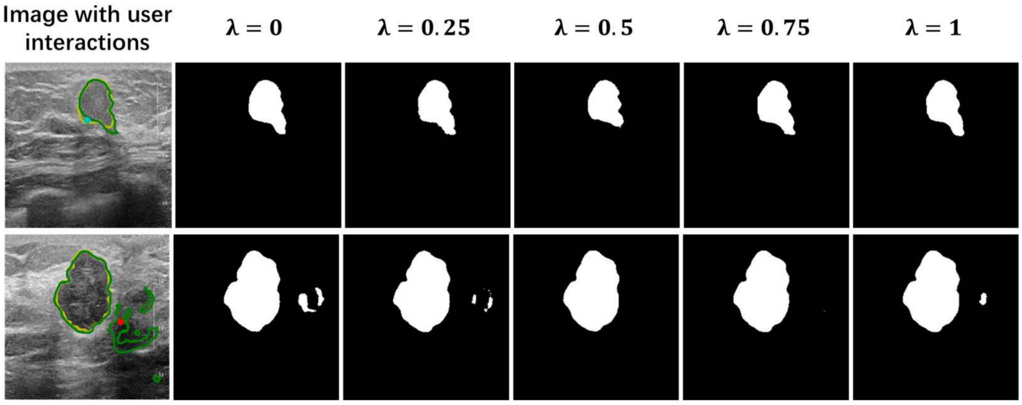

Figure 6. The segmentation results of WDTISeg having different values of λ by given the same user interaction points.

The combination of Euclidean distance transform and geodesic distance transform can both convey the location information of interaction points and make use of image context information. The experimental results demonstrate that this combination improves the stability of the segmentation model to cope with images that are difficult to segment.

3.5. Effect of Proposed Loss Function

We explored the effect of our involved loss function by observing the dice rate on the training and validation datasets, as shown in Figure 7. On the plot of dice rate on the training dataset, $\mathcal{L}_{\text{Dice}}$(Dice loss) achieved the best performance, while $\mathcal{L}_{\text{total}}$ (BCE + Dice +

SC loss) came second. The reason why $\mathcal{L}_{\text{Dice}}$ performed well on the training set is that the network used dice as the evaluation metric, and the network maximizes dice by optimizing the network structure during training. However, the dice rate of $\mathcal{L}_{\text{Dice}}$ on the validation dataset had a sharp oscillation. This is mainly because $\mathcal{L}_{\text{Dice}}$ is a region-dependent loss, and if some pixels of a small target are incorrectly predicted, then it will lead to a significant change in the loss value, which will result in a drastic change in the gradient.

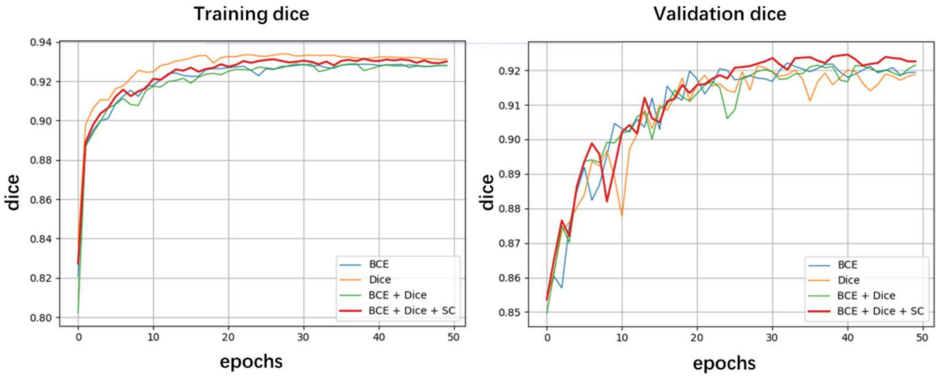

Figure 7. Dice rate on the training (left) and validation datasets (right) of different loss functions.

Compared with other three loss functions, $\mathcal{L}_{\text{total}}$ incorporating prior knowledge achieves optimality on the validation set and does not show more intense oscillations after epochs greater than 30. The BCE loss used for dichotomous classification is insensitive to category imbalance, so it can prevent the oscillation due to $\mathcal{L}_{\text{Dice}}$ to some extent. On the other hand, the loss function $\mathcal{L}_{\text{SC}}$ based on compact shape constraint utilizes prior information of tumor shape, and thus can improve the accuracy of segmentation. Note that $\mathcal{L}_{\text{SC}}$ converges to 1 at the minimum when the tumor is circular, so it cannot be used as a segmentation loss function alone.

The purpose of introducing a subjective human into the segmentation process is to use human's prior knowledge as a supplement to improve the segmentation accuracy. In the interactive segmentation task, human is both the participant in the interactive segmentation process, the prior information provider, and also the evaluator of segmentation results without the ground truth. In the interaction segmentation task, we may learn from the few-shot learning which has been widely used machine learning classification task. Human guides segmentation on a few simple images so that the network can master the segmentation skills, further reducing the training time and human interaction time.

3.6. Quantitative Comparison of Different Methods

We evaluated WDTISeg with graph cuts, random walker, and DeepIGeoS (R-net). Table 1 presents a quantitative comparison of these methods on the testing data. All results are accepted after 8 interactions for the interactive segmentation method. Compared with the other three methods, the dice of WDTISeg ($\lambda = 0.5$, $\mathcal{L}_{\text{total}}$) reached 94.45 ± 3.26 (%) after 8 interactions, which fully shows that our method can achieve a high segmentation accuracy with fewer interactions.

Visual comparison results are shown in Figure 8. All interactive segmentation methods can attain a high accurate segmentation after enough interactions. However, results of graph cuts and random walker showed more rough edges. In contrast, our method was able to obtain a segmentation that fit more closely to the tumor margin. What is more, our WDTISeg only required simple point clicks, while graph cuts and GrabCut require more scribbles or a bounding box. The results of DeepIGeoS are more similar to that of our method, because we also used distance conversion to pass interaction information. However, it can be found that the segmentation result of our method was smoother at the tumor edge, especially at the lower right corner of case 4. This may benefit from the fact that we used a shape constraint loss to impose prior constraints on tumor shape.

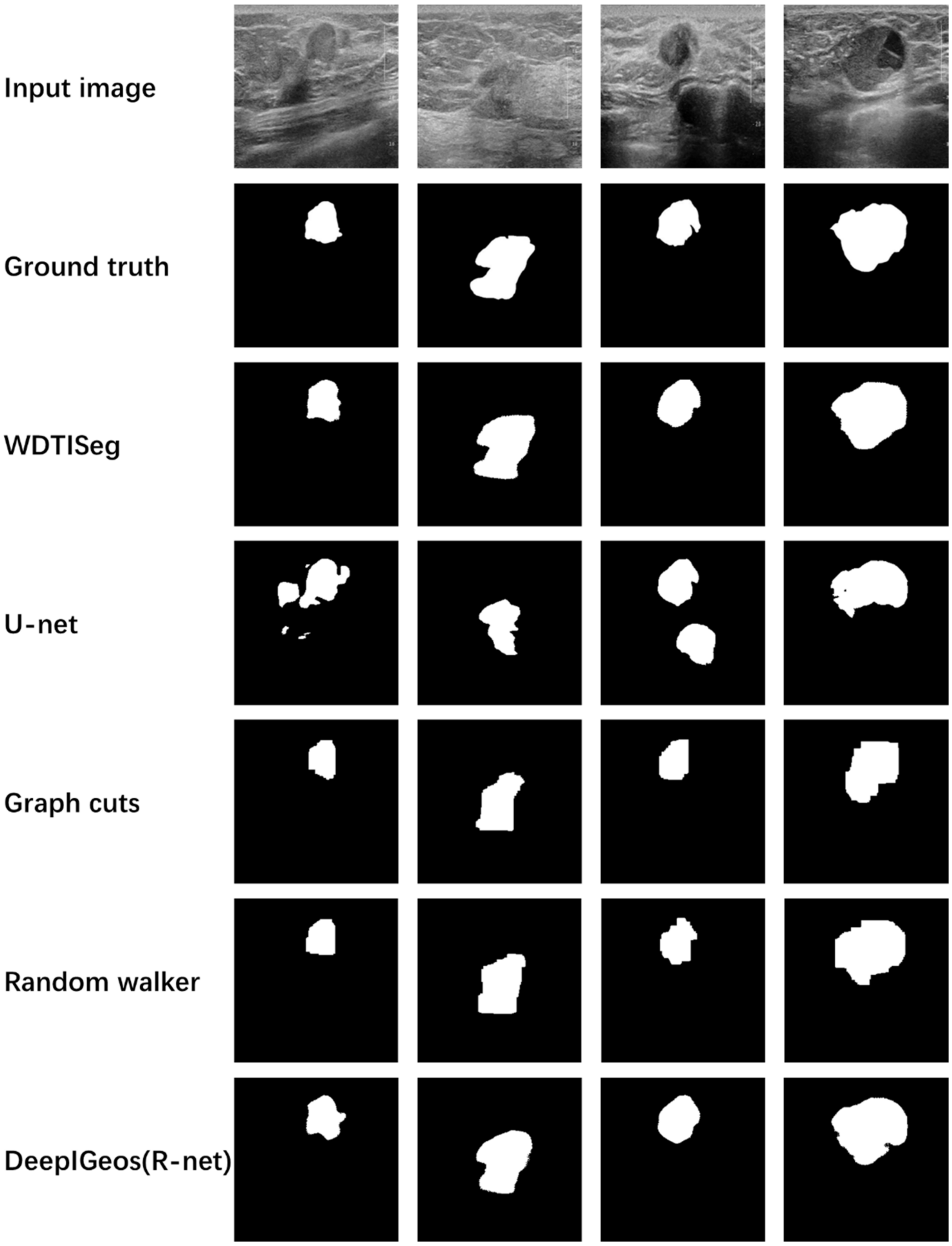

Figure 8. Visual comparison for breast ultrasound image segmentation. All results were the final result accepted after 8 interactions.

4. Conclusions

In this paper, we proposed a one-stage interactive segmentation framework (WDTISeg) for breast ultrasound image segmentation. The ultrasound image was put into the network first to attain an initial segmentation, on which user interaction points were provided to indicate mis-segmentations. Interaction points were converted into distance maps by weighted distance transform to be part of input of the interactive network. The one-stage network of point interaction made the interaction simpler. The loss function designed for the clinical prior knowledge of breast cancer further improved the segmentation accuracy. Comparison with other methods on the test dataset demonstrated the advantages of the proposed method.

However, our method had limitations in combining the two distance transforms. In this paper, in order to verify the usefulness of combining two distance conversion methods, different ratios were tried to conduct experiments, and the experimental results proved in general that combining two methods helps to improve the segmentation accuracy. Considering the differences of different ultrasound images, the most suitable combination ratio should be different for each image. If an optimal ratio value can be obtained adaptively according to the characteristics of the ultrasound image itself, thus attaining the best segmentation result, it will further improve the segmentation accuracy and enhance the segmentation robustness.

Author Contributions: Conceptualization, X.L., M.Q. and Y.G.; methodology and experiments, X.L.; data curation, J.Z., S.Z. and C.C.; writing and editing, X.L. and Y.G.; supervision, Y.G. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant 61871135, 81627804 and 81830058). The asterisk indicates the corresponding author.

Institutional Review Board Statement: The retrospective study was approved by the institutional ethics committee of Fudan University Shanghai Cancer Center, and the informed consent was waived.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cheng, H.D.; Shi, X.J.; Min, R.; Hu, L.M.; Cai, X.P.; Du, H.N. Approaches for Automated Detection and Classification of Masses in Mammograms. *Pattern Recognit.* **2006**, *39*, 646–668. [CrossRef]
2. Monticciolo, D.L.; Newell, M.S.; Moy, L.; Niell, B.; Monsees, B.; Sickles, E.A. Breast Cancer Screening in Women at Higher-Than-Average Risk: Recommendations From the ACR. *J. Am. Coll. Radiol.* **2018**, *15*, 408–414. [CrossRef] [PubMed]
3. Sharma, N.; Ray, A.; Shukla, K.; Sharma, S.; Pradhan, S.; Srivastva, A.; Aggarwal, L. Automated Medical Image Segmentation Techniques. *J. Med. Phys.* **2010**, *35*, 3. [CrossRef] [PubMed]
4. Ramadan, H.; Lachqar, C.; Tairi, H. A Survey of Recent Interactive Image Segmentation Methods. *Comp. Vis. Media* **2020**, *6*, 355–384. [CrossRef]
5. Xu, N.; Price, B.; Cohen, S.; Yang, J.; Huang, T. Deep Interactive Object Selection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 373–381.
6. Wang, G.; Li, W.; Zuluaga, M.A.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S.; et al. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Trans. Med. Imaging* **2018**, *37*, 1562–1573. [CrossRef] [PubMed]
7. Wang, G.; Zuluaga, M.A.; Li, W.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S.; et al. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1559–1572. [CrossRef] [PubMed]
8. Lin, Z.; Zhang, Z.; Chen, L.-Z.; Cheng, M.-M.; Lu, S.-P. Interactive Image Segmentation with First Click Attention. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 13336–13345.
9. Zhang, J.; Shi, Y.; Sun, J.; Wang, L.; Zhou, L.; Gao, Y.; Shen, D. Interactive Medical Image Segmentation via a Point-Based Interaction. *Artif. Intell. Med.* **2021**, *111*, 101998. [CrossRef] [PubMed]
10. Boykov, Y.Y.; Jolly, M.-P. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. In Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; Volume 1, pp. 105–112.
11. Rother, C.; Kolmogorov, V.; Blake, A. GrabCut -Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Trans. Graph. (SIGGRAPH)* **2004**, *23*, 309–314. [CrossRef]
12. Grady, L.; Schiwietz, T.; Aharon, S.; Westermann, R. Random Walks for Interactive Organ Segmentation in Two and Three Dimensions: Implementation and Validation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005, Palm Springs, CA, USA, 26 October 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 773–780.
13. Bai, X.; Sapiro, G. A Geodesic Framework for Fast Interactive Image and Video Segmentation and Matting. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
14. Criminisi, A.; Sharp, T.; Blake, A. GeoS: Geodesic Image Segmentation. In *Computer Vision—ECCV 2008*; Forsyth, D., Torr, P., Zisserman, A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5302, pp. 99–112, ISBN 978-3-540-88681-5.
15. Price, B.L.; Morse, B.; Cohen, S. Geodesic Graph Cut for Interactive Image Segmentation. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3161–3168.

16. Gulshan, V.; Rother, C.; Criminisi, A.; Blake, A.; Zisserman, A. Geodesic Star Convexity for Interactive Image Segmentation. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3129–3136.
17. Zhang, S.; Liew, J.H.; Wei, Y.; Wei, S.; Zhao, Y. Interactive Object Segmentation with Inside-Outside Guidance. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 12231–12241.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241, ISBN 978-3-319-24573-7.
19. Wu, Y.; He, K. Group Normalization. In *Computer Vision–ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 11217, pp. 3–19, ISBN 978-3-030-01260-1.
20. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6 July 2015; Volume 35, pp. 448–456.
21. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv* **2015**, arXiv:1505.00853.
22. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Madison, WI, USA, 21 June 2010; pp. 807–814.
23. Dijkstra, E.W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1*, 269–271. [CrossRef]
24. Sethian, J.A.; Vladimirsky, A. Fast Methods for the Eikonal and Related Hamilton- Jacobi Equations on Unstructured Meshes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5699–5703. [CrossRef] [PubMed]
25. Nosrati, M.S.; Hamarneh, G. Incorporating Prior Knowledge in Medical Image Segmentation: A Survey. *arXiv* **2016**, arXiv:1607.01092.
26. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.